Cross-entropy Analysis of the Information in Forensic Speaker Recognition

Daniel Ramos *and Joaquin Gonzalez-Rodriguez*
Escuela Politecnica Superior. Universidad Autonoma de Madrid
daniel.ramos@uam.es
http://atvs.ii.uam.es

Outline

- Assessment and admissibility in forensic science
- Information-theoretical assessment of Likelihood Ratio values
  - The *ECE* plot
- Experimental example
- Simulated forensic speaker recognition case
- Conclusions

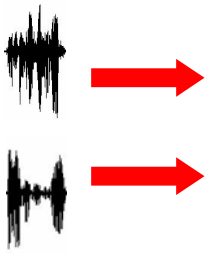
Assessment and admissibility in forensic science

Admissibility of evidence

- Example: American *Daubert* rules -1993-
- Admissibility considers:
  - Empirical test: falsifiable / repeatable
    - Not only in the lab, but also in real-case conditions
  - Known accuracy (e.g., error rates)
  - Peer-reviewed and published
  - Standards define its use
  - General acceptance among the community
- Clear needs
  - Testability
  - Assessment of techniques used for forensic evidence evaluation

The question

- We compute a Likelihood Ratio (LR) using forensic speaker recognition



Forensic Speaker Recognition

$$LR = \frac{p\left(E\middle|\theta_p, I\right)}{p\left(E\middle|\theta_d, I\right)}$$

- How accurate is my LR?
  - Admissibility condition: assessment of the accuracy
  - …but in which terms?

Information-theoretical assessment of LR values

Accuracy of the LR

- Recently proposed measures of accuracy of LR values
  - C_{llr} [Brümmer 2006]
    - Cost-based, information-theoretical
    - But assumes prior equals 0.5
      - Controversial in LR-based forensic identification
  - Normalized Cross-Entropy (NCE) [Campbell 2005]
    - Information-theoretical
    - But does not clearly separate the contribution of the prior and the LR
      - Controversial in LR-based forensic identification
- Sources of controversy
  - Priors are province of the fact finder
  - Priors are dependent on each given case

Information-theoretical assessment of the accuracy of the LR

- Accuracy of the LR: Empirical Cross-Entropy (ECE)
  - According to previous approaches
    - Normalized Cross-Entropy is basically a normalized version of ECE
    - C_{llr} is the value of ECE assuming prior equals 0.5
- Novel representation (ECE plot)
  - Average information needed for obtaining the true value of the hypothesis in a case:
    - θ_p: suspect is the author of the questioned recording
    - θ_d: another individual is the author of the questioned recording
  - It keeps the separation of roles among scientist and fact finder
    - The influence of the prior is separated from the LR

Entropy

- Expected uncertainty of a random variable
- Example: correct hypothesis

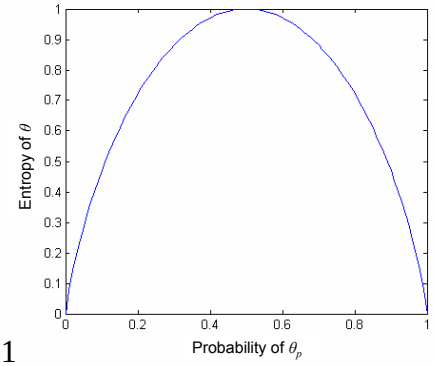
$$\theta = \{\theta_p, \theta_d\}$$

- Zero if certainty

$$P(\theta_i) = 1$$

$$P(\theta_j) = 0 \quad (j \neq i)$$

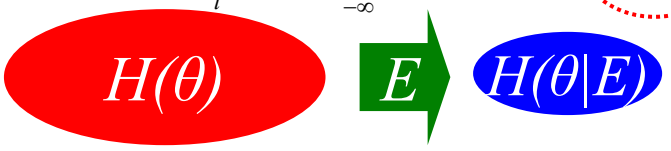
- Maximum for maximum uncertainty $P(\theta_p) = P(\theta_d) = \frac{1}{2}$
- Measured in *bits* (for base-2 logarithms)



Expected entropy after knowing E

- Conditional entropy: expected uncertainty after knowledge of E

$$H_P(\theta|E) = -\sum_i P(\theta_i) \int_{-\infty}^{\infty} p(e|\theta_i) \log_2 P(\theta_i|e)\, de$$



- Information is defined as the expected reduction of uncertainty
- Conditional entropy is difficult to compute in general
  - Which posterior probability to use...?

Cross-entropy

- Defined as:

$$H_{\tilde{P}\|P}(\theta|E) = -\sum_i P(\theta_i) \int_{-\infty}^{\infty} \tilde{p}(e|\theta_i) \log_2 P(\theta_i|e)\, de$$

- Two distributions
  1. Posterior distribution obtained using the LR of the forensic system
     - ...and the prior from the fact finder

$$P(\theta_p|e) = \frac{LR \cdot \frac{P(\theta_p)}{P(\theta_d)}}{1 + LR \cdot \frac{P(\theta_p)}{P(\theta_d)}}$$

  2. A reference probability

Empirical Cross-Entropy (*ECE*)

- Empirical method to measure cross-entropy
  - From a speech evaluation database:
    - Target LR values (θ_p is true) from scores $\mathbf{E}_p$
    - Non-target LR values (θ_d is true) from scores $\mathbf{E}_d$
- *ECE* is computed empirically (average as expectation)

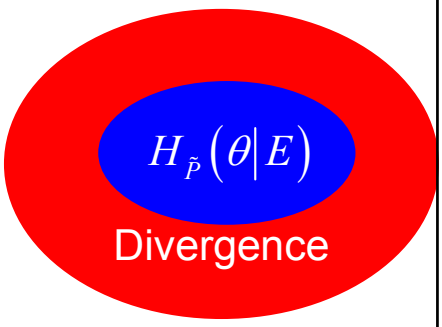
$$H_{\tilde{P}\|P}(\theta|E) \approx ECE = -P(\theta_p)\frac{1}{N_p}\sum_{j\in\mathbf{E}_p} \log_2 P(\theta_p|e_j) - P(\theta_d)\frac{1}{N_d}\sum_{j\in\mathbf{E}_d} \log_2 P(\theta_d|e_j)$$

ECE interpretation

- Cross entropy is decomposed as:

$$ECE \simeq H_{\tilde{P}\|P}\left(\theta\middle|E\right) = H_{\tilde{P}}\left(\theta\middle|E\right) + D_{\tilde{P}\|P}\left(\theta\middle|E\right)$$

- Entropy of the reference
  - Uncertainty if the reference is used
- Divergence
  - from the system's posterior
  - w.r.t. the reference
  - Information loss
    - Because we expect the reference and the we obtain the system's LR



Choosing an intuitive reference

- We propose to choose the following reference:

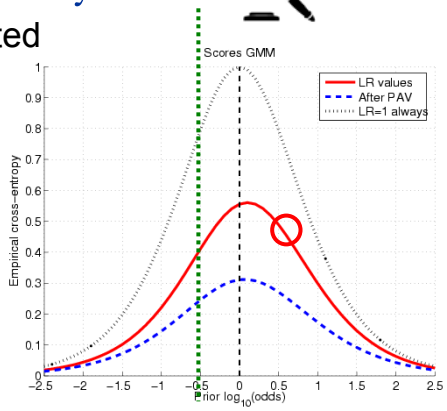
$$\tilde{P}\left(\theta_p\middle|e\right) = 1 \quad \theta_p \text{ true}$$

$$\tilde{P}\left(\theta_p\middle|e\right) = 0 \quad \theta_d \text{ false}$$

  - "…as if the fact finder would know the true answer"
  - The reference is "certainty"
- The entropy of this reference is zero
  - Cross-entropy: divergence of the system's LR values from "certainty"

ECE plots: LR accuracy

- Three systems are represented
  - System's LR values (solid)
  - Always LR=1 (dotted)
  - Calibrated LR values (dashed)
    - True answers are needed
    - Pool Adjacent Violators (PAV) algorithm [Brümmer 2006]
  - C_{llr}: ECE at prior 0.5

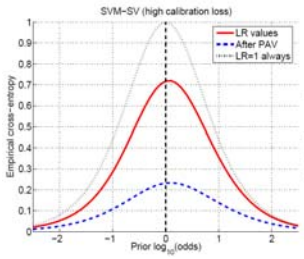


- *ECE* also measured in bits
- Separation of roles
  - Forensic scientist: *ECE* computation for a wide range of priors
    - Because the scientist cannot set the prior…
  - Fact finder: prior establishment and measure of *ECE* in the plot

Experimental comparison

Comparison of LR computation techniques

- ATVS systems, NIST SRE 2006 protocol



SVM-SV system	GMM system	Fused system
T-normed scores	T-normed scores	Logistic regression
		(LR computation)
Good discrimination loss	Good discrimination loss	Good discrimination loss
(low ECE after PAV)	(low ECE after PAV)	(low ECE after PAV)
High ECE due to calibration loss	High ECE due to calibration loss	Good calibration loss
		(ECE ≈ ECE after PAV)

Simulated case

Simulated Case

- Incriminating recordings wire-tapped by police
- The population of potential offenders is the population of Madrid
  - Prior of 1 over 5 million people?

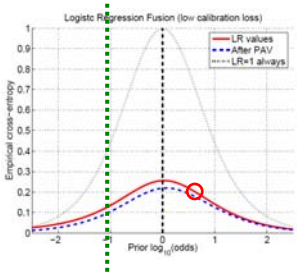
- There is other prior information
  - Police investigations (witnesses, other evidence, etc.) reduce the list of suspects to 11 people
    - Equally likely to be the author

Role of the Fact Finder

- A suspect is selected among the 11 potential offenders

- The fact finder sets the prior:
  - Equally likely
  - Thus, probability of 1 over 11 to be the author
  - The prior may be unknown by the forensic scientist
- The fact finder asks the forensic scientist:
  - To evaluate the evidence
  - To assess system accuracy

Role of the Forensic Scientist



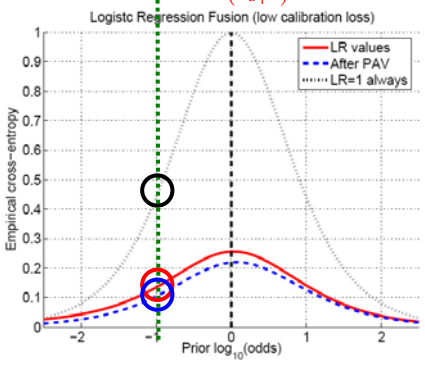
- Assess the accuracy of the system
  - Priors may be unknown
  - Priors change case by case
  - Using *ECE* plots...
    - prior is not stated
    - accuracy computed at any prior
- After system validation in conditions matched to the case
  - Compute and report the LR
  - Fact finder can infer a posterior
    - From LR and prior

- Once prior is stated by the fact finder
  - *ECE* (accuracy) can be computed

Case-based *ECE* plot interpretation

$$P(\theta_p|I)=\frac{1}{11}$$

$$\Rightarrow \frac{P(\theta_p|I)}{P(\theta_d|I)}=\frac{1}{10}$$



- Before analyzing the evidence, I would need a great amount of information (0.45 bits) in order to know whether the suspect is the author of the questioned recording or not
- After evidence analysis, the amount of information needed is significantly smaller (0.12 bits)
- With perfect calibration, I would need almost the same information as the system (0.1 bits)
  - This performance requires the true answers
- System is validated for every prior
  - Thus, the LR value yielded by the system is useful

Conclusions

Conclusions

- Upcoming admissibility requirements in forensic science demand the assessment of forensic disciplines
- A measure of accuracy (ECE) has been proposed in terms of information-theoretical quantities
  - According to previous works in the literature (C_{llr}, NCE)
- We present ECE plots as a way to represent accuracy of a set of LR values
  - It integrates previous approaches
  - It preserves competences of fact finders
  - It has an intuitive interpretation
- The technique has been illustrated by
  - Experimental results (NIST SRE 2006)
  - A simulated forensic speaker recognition case
- We have recently assessed other forensic disciplines with ECE plots
  - Glass and paint evidence analysis [Ramos et al. 2007]
  - LR values obtained by different techniques [Aitken et al. 2007]